Aquat. Living Resour. 18, 159–168 (2005)

DOI: 10.1051/alr:2005017
www.edpsciences.org/alr

Aquatic
Living
Resources

Species richness estimators applied to fish in a small tropical river sampled by conventional methods and rotenone

Łukasz Głowacki[a] and Tadeusz Penczak

Department of Ecology and Vertebrate Zoology, University of Łódź, 12/16 Banacha Str., 90-237 Łódź, Poland

Received 7 June 2004; Accepted 25 April 2005

Abstract – A number of analytical tools were applied to estimate fish species richness in a small tropical stream (Brazil) that had been sampled by conventional methods (double stick net, electrofishing) and rotenone. Three sites were compared in two seasons each and a total of 34 species were captured. The sampling efficiency of electrofishing was much better than that of the double stick net and little worse than of rotenone. No estimator was thoroughly consistent and accurate in predicting species richness on the basis of electrofishing, but the Homogenous Model, which assumes that all species have the same detection probabilities, was best. The rarefaction technique should be used with caution as it may produce misleading results in species richness predictions.

Key words: Freshwater fish / Electrofishing / Rotenone / Species richness estimators / Brazil

Résumé – Estimateurs de la richesse en espèces appliqués aux poissons d'un cours d'eau tropical avec des modes de capture conventionnels et au moyen de roténone. Plusieurs outils analytiques ont été appliqués pour estimer la richesse en espèces de poissons d'un cours d'eau tropical (Brésil), l'échantillonnage a été effectué au moyen de méthodes conventionnelles (épuisette, pêche électrique) et en utilisant de la roténone. Trois sites ont été comparés lors de chacune des deux saisons, et un total de 34 espèces ont été capturées. L'efficacité de l'échantillonnage par pêche électrique est meilleure que celle effectuée par épuisette mais légèrement moindre que celle effectuée à la roténone. Il n'y a pas d'estimateurs cohérents et précis pour la prédiction du nombre d'espèces, basée sur la pêche électrique. Cependant, le modèle homogène, qui suppose que toutes les espèces ont les mêmes probabilités d'être détectées, est le meilleur. La technique de raréfaction doit être utilisée avec précaution car elle peut produire des résultats erronés dans la prédiction de la richesse en espèces.

1 Introduction

Reliable qualitative sampling is one of main technical objectives of fishery research when species richness and assemblage structure are to be determined (Matthews 1998; Jackson et al. 2001). It aims to obtain the actual number of species in a study area. Bayley and Peterson (2001), Peterson and Rabeni (2001), Hughes et al. (2002), Meador et al. (2003) and Reynolds et al. (2003) lately undertook extensive sampling and research to estimate fish species number in USA rivers. Their effort was directed to extend the sampled area or length of sampled reach, and to test numerous theoretical techniques available in literature to search for the best model to analyze obtained data. In our investigations in a tropical stream we had a control (rotenone), which is considered a more objective gear of obtaining reliable qualitative and quantitative estimates of fish faunas by most fish scientists (Davies and Shelton 1983). Hence, perhaps, our data may be more reliable and objective for model validations.

[a] Corresponding author: glowacki@biol.uni.lodz.pl

Qualitative sampling is hardly ever perfect. One reason is that fish of some species escape being captured (in the present study for example). Another one, for example, is seasonal movement of some (even non-migratory but obligatory riverine) species: they periodically enter different reaches of given streams or move to other rivers of a given river system (Penczak and Jakubowski 1990; Matthews 1998).

To maximize qualitative sampling efficiency various approaches are used. The dominant approach to overcome mostly the first above limitation is simultaneous application of several catching methods; an approach to overcome also the second are indirect, deduction techniques that compensate for species being missed (Matthews 1998; Bayley and Peterson 2001; Peterson and Rabeni 2001; Hughes et al. 2002; Reynolds et al. 2003).

Obtaining the reliable species number is of high theoretical and practical importance: more theoretical on larger scales and more practical on smaller scales (Matthews 1998). The former because data from various regions of the earth enable detecting global diversity patterns (Oberdorff et al. 1995;

Matthews 1998; Meador et al. 2003), while the latter because they enable application of more effective measures for preservation, restitution and/or exploitation of biotic resources (Palmer 1994, 1995; Matthews 1998; Jackson et al. 2001).

The present study contributes to solving the above mentioned problems: on the one hand it supplies data on a habitat and its fish assemblage that are much different from European or North American ones and on which very few reports exist, on the other it analyses the data with a number of indirect techniques to evaluate the efficiency of qualitative fishing. It follows a former study (Penczak et al. 2003) that described the density of the assemblage of the stream with known error and calibrated the two fishing conventional gears employed (double stick net and electrofishing) with rotenone.

However, in contrast to the former study (Penczak et al. 2003), which mostly analysed the problem of mean fish weight from the viewpoint of bioenergetics models and of ecological/spawning groups of fish, the present study deals with the problem of how sample pooling affects the performance of species richness estimators.

The "performance" should be considered in relative terms, because each estimator assumes some biotic and abiotic set of conditions of the sampled animal assemblage, which are either actually met or not met. Consequently, it is equally justified to speak of a "performance" of the small Brazilian stream fish fauna in terms of these estimators as well.

2 Material and methods

2.1 Sampling

Sampling was carried out in December and March in three contiguous sites established in the lower course of the Taquaral Stream, which empties into the Corumba River about 15 km upstream of the backwater of a newly created, large, man-made reservoir (Goiás State, subtropical Brazil) (Penczak et al. 2003). The sites were 40 m long each, about 2 m wide and 0.2 m deep. Site 3 much differed from the others: it was not canopied (70% canopy in sites 1 and 2), and much more stony and much less sandy, while it lacked overhanging shrub branches and snags that were common in sites 1 and 2. Discharge was, on average, by half higher in March (27 and 18 $\mathrm{L\,s^{-1}}$, respectively) and conductivity was by about 25% lower in March (ca. 40 versus 30 $\mu\mathrm{S\,cm^{-1}}$) in all the sites. However, for the sampling purpose conductivity was increased several times by salting water (Penczak et al. 2003). Site 1 was first sampled with a double stick net (DSN) and sites 2 and 3 with electricfishing (EF). Rotenone (R) was then pumped just upstream of the blocking net of site 3, the most upstream one. Rotenone dilution was calculated after Davies and Shelton (1983) to collect uncaptured fish. Details of sites morphology, map, other environmental conditions, and sampling techniques were presented in Penczak et al. (2003).

2.2 Analytical methods

Data were first analysed using the rarefaction technique (Hurlbert 1971; Heck et al. 1975; Ludwig and Reynolds 1988; Magurran 1988), which has been successfully applied to fisheries research (Penczak et al. 1998; Głowacki and Penczak 2000). The shape of the rarefaction analysis curve is presumed to indicate whether the list of captured species is complete. This should happen when the rarefaction curve levels off.

EstimateS software (Version 5, R.K. Colwell 1997 `http://viceroy.eeb.uconn.edu/estimates`) was then applied to do pair analysis. Later on, we applied 10 well-known species richness estimators using the SPADE (Chao and Shen 2003) program to our data; the estimators used were (formulas of only those that were more successful are presented; the cut-off point for rare species was always 10):

Homogenous Model (HM) (Chao and Lee 1992; Chao et al. 2000):

$$\hat{S} = D_{\mathrm{abun}} + \frac{D_{\mathrm{rare}}}{\hat{C}_{\mathrm{rare}}}, \quad \text{where}$$

$D_{\mathrm{rare}} = \sum_{i=1}^{\kappa} f_i$, $\hat{C}_{\mathrm{rare}} = 1 - f_i / \sum_{i=1}^{\kappa} i f_i$ and κ = cut-off point.

Homogenous Model (Maximum Likelihood Estimation, MLE) (HM (MLE)) (Chao and Lee 1992):

$$\hat{S} \quad \text{is the solution of} \quad D = S\left[1 - \exp(-n/S)\right]$$

Chao 1 (Chao 1984; Shen et al. 2003):

$$\hat{S} = \begin{cases} D + f_i^2/(2f_2), & \text{if} \quad f_2 > 0 \\ D + f_1(f_1 - 1)/2, & \text{if} \quad f_2 > 0 \end{cases}$$

ACE (Chao and Lee 1992; Chao et al. 2000):

$$\hat{S} = D_{\mathrm{abun}} + \frac{D_{\mathrm{rare}}}{\hat{C}_{\mathrm{rare}}} + \frac{f_i}{\hat{C}_{\mathrm{rare}}}\hat{\gamma}^2_{\mathrm{rare}}$$

ACE 1 (Chao and Lee 1992):

$$\hat{S} = D_{\mathrm{abun}} + \frac{D_{\mathrm{rare}}}{\hat{C}_{\mathrm{rare}}} + \frac{f_i}{\hat{C}_{\mathrm{rare}}}\tilde{\gamma}^2_{\mathrm{rare}}$$

Jackknife 1 (Burnham and Overton 1978):

$$\hat{S} = D + \frac{n-1}{n} f_1$$

Jackknife 2 (Burnham and Overton 1978):

$$\hat{S} = D + \frac{2n-3}{n} f_1 - \frac{(n-2)^2}{n(n-1)} f_2$$

Gamma-Poisson Model (G-P M), Gamma-Poisson UMLE (G-P UMLE) and Gamma-Poisson CMLE (G-P CMLE) (Chao and Bunge 2002):

The notation used in the above formulas is the following:

- S total number of species in a community;
- X_i number of individuals (frequency) the ith species is observed in the sample, $I = 1, 2, \ldots, S$ (only species with X_i are observable in the sample);
- n sample size, $n = \Sigma_{i=1}^{S} Xi = \Sigma_{j\geq 1} j f_j$. ($f$'s are defined below);
- $I[A]$ the usual indicator function, i.e., $I[A] = 1$ if the event A occurs, 0 otherwise.

f_j number of species that are represented exactly j times in the sample, $j = 0, 1, \ldots, n$, $f_j = \Sigma_{i=1}^{S} I[X_i = j]$. ($f_0$ denotes the number of unobserved species).

C sample coverage;

$\hat{}$ an estimator from data, e.g., $\hat{S}$ (estimator of S) and $\hat{C}$ (estimator of C) … etc.

D number of distinct species discovered in the sample, ($D = \Sigma_{i=1}^{S} I[X_i > 0] = \Sigma_{j\geq 1} f_j$);

κ the cut-off point (default = 10), which separates species into "abundant" and "rare" groups for abundance data; it separates species into "frequent" and "infrequent" groups for incidence data;

D_{rare} the number of distinct species for "rare" group, $D_{\text{rare}} = \Sigma_{i=1}^{\kappa} f_i$.

D_{abun} the number of distinct species for "abundant" group, $D_{\text{rare}} = \Sigma_{i>\kappa}^{\kappa} f_i$.

$\hat{C}_{\text{rare}}$ estimated sample coverage for "rare" group, $\hat{C}_{\text{rare}} = 1 - f_1/\Sigma_{i>\kappa}^{\kappa} f_i$.

$\hat{\gamma}_{\text{rare}}$ estimated coefficient of variation;

$\widetilde{\gamma}_{\text{rare}}$ estimated coefficient of variation for a highly heterogeneous case.

The performance of these estimators was assessed using Palmer's (1990, 1991) categories of bias and precision. Bias, or deviation from an ideal estimator, is the ratio of over- to underestimates (an ideal estimator would overestimate true values 50% of the time, when its value is close to 100% or 0% it is very biased), while precision is closeness to true values (in our case measured in percentages).We added a third measure, the correlation between a given estimator and numbers of actually captured species in samples and their pools. Correlation and other statistical analysis were performed with Statistica, ver. 5.5 A (StatSoft, Inc. 2000) computer software. Figures were also prepared using the Statistica.

2.3 Data structures

In the Taquaral Stream, Penczak et al. (2003) sampled 2−3 times with DSN in site 1, while 3−5 times with EF in sites 2 and 3 and once with rotenone in each site on each sampling occasion. As a result, they obtained elementary raw samples (Table 1), which are marked using the following symbol system: the first literal stands for season (December or March), numeral following the literal for a site (1, 2 or 3), subsequent 1−3 literals for a sampling gear (DSN, EF, or Rotenone) and the last numeral for the ordinal number of a pass with a given gear.

These elementary samples were then totaled by gear producing 12 samples (Tables 1 and 2).

Note that in further analysis (Table 3) sample pools were species numbers obtained by EF (column *CT* in Tables 3 and 4). Species numbers obtained by the control were those that were captured by rotenone application after each EF application. But the "*R* only" (Table 3) numbers were obtained by comparing a given EF pool with its respective rotenone pool, not by comparing each sample of an EF pool with its respective rotenone sample and then summing the species numbers up. For example, in the case of pool 4+5+6 we compared it with the 10+11+12 pool and not 4 with 10, 5 with 11 and 6 with 12 to sum them up, which would produce a different estimate.

Table 1. Taquaral sample pooling/coding pattern. Initial D signifies December, initial M signifies March, numeral following them is site number. DSN: double stick net, EF: electrofishing, R: rotenone.

Initial sample	(Pooled and) coded as
D1DSN1, D1DSN2	(1) D1DSN
M1DSN1, M1DSN2, M1DSN3	(2) M1DSN
D2EF1, D2EF2, D2EF3, D2EF4	(3) D2EF
M2EF1, M2EF2, M3EF3, M4EF4	(4) M2EF
D3EF1, D3EF2, D3EF3, D3EF4, D3EF5	(5) D3EF
M3EF1, M3EF2, M3EF3, M3EF4, M3EF5	(6) M3EF
D1R	(7) D1R
M1R	(8) M1R
D2R	(9) D2R
M2R	(10) M2R
D3R	(11) D3R
M3R	(12) M3R

3 Results

The curves that rarefaction produced are grouped in two figures: one profile (Figs. 1a–f) compares seasonal differences in samples obtained by the same gear (in each subfigure), while the other compares, in each subfigure (Figs. 2a–i), site differences produced by different DSN/EF and rotenone samples.

The curve leveled-off to a higher degree in December than in March and more so sites 2 and 3 than 1 (Fig. 1). In site 1 both DSN and rotenone curves little leveled-off, but more so in December (Figs. 1a,b). The same, but to a higher degree occurred in sites 2 and 3 (Figs. 1c–f). A similar pattern is obtained when seasons (December and March) are compared (Figs. 2a–i).

Consequently, the rarefaction technique indicated that the list of species (as obtained by DSN, EF or rotenone) was most complete in samples 3 and 5, less in 4 and 6 and least in 1 and 2. This might suggest that the gears varied in sampling efficiency due to their selectiveness.

One method of determining whether it might be true was an attempt to inspect common species occurring in sample pairs Some of the total of 66 possible sample pairs are presented in Table 2. Covariance, for example, calculated between given DSN samples and all the rotenone samples (Tables 2B-C) revealed a difference between them and rotenone ones. While the March DSN sample covaried with rotenone almost as much as EF samples (9.1 and 8.6, respectively), the December DSN sample displayed negative covariance (−0.8), which would suggest that DSN was much selective. On the other hand a comparison of given EF samples with their respective rotenone samples (Table 2A) provided much stronger covariance between rotenone and common species (12.3) than between the EF and common species (2.3).

A more general observation was that EF samples of site 3 (samples 5 and 6 in Tables 2F and G) considerably worse covaried with rotenone samples (5.36 and 4.4) than EF samples of site 2) (i.e. numbered 3 and 4 in Tables 2D and E) with rotenone samples (8.6 and 9.1, respectively). We also wanted

Table 2. Sample pairs obtained in the Taquaral River by DSN (double stick net), EF (electrofishing) or Rotenone (R) and covariance calculated between them. Initial D signifies December, initial M signifies March; numeral following them is site number.

First sample	Second sample	Number of species in the first sample	Number of common species	Number of species in the second sample
A				
			Covariance = 12.3	
3 (D2EF)	9 (D2R)	19	15	17
4 (M2EF)	10 (M2R)	21	15	17
5 (D3EF)	11 (D3R)	19	8	9
6 (M3EF)	12 (M3R)	18	9	10
		Covariance = 2.3		
B				
1 (D1DSN)	7 (D1R)	12	9	16
1 (D1DSN)	8 (M1R)	12	5	16
1 (D1DSN)	9 (D2R)	12	6	17
1 (D1DSN)	10 (M2R)	12	6	17
1 (D1DSN)	11 (D3R)	12	6	9
1 (D1DSN)	12 (M3R)	12	7	10
			Covariance = −0.8	
C				
2 (M1DSN)	7 (D1R)	12	9	16
2 (M1DSN)	8 (M1R)	12	8	16
2 (M1DSN)	9 (D2R)	12	6	17
2 (M1DSN)	10 (M2R)	12	10	17
2 (M1DSN)	11 (D3R)	12	4	9
2 (M1DSN)	12 (M3R)	12	7	10
			Covariance = 4.4	
D				
3 (D2EF)	7 (D1R)	19	15	16
3 (D2EF)	8 (M1R)	19	12	16
3 (D2EF)	9 (D2R)	19	15	17
3 (D2EF)	10 (M2R)	19	12	17
3 (D2EF)	11 (D3R)	19	8	9
3 (D2EF)	12 (M3R)	19	8	10
			Covariance = 8.6	
E				
4 (M2EF)	7 (D1R)	21	13	16
4 (M2EF)	8 (M1R)	21	13	16
4 (M2EF)	9 (D2R)	21	12	17
4 (M2EF)	10 (M2R)	21	15	17
4 (M2EF)	11 (D3R)	21	7	9
4 (M2EF)	12 (M3R)	21	8	10
			Covariance = 9.1	
F				
5 (D3EF)	7 (D1R)	19	15	16
5 (D3EF)	8 (M1R)	19	10	16
5 (D3EF)	9 (D2R)	19	12	17
5 (D3EF)	10 (M2R)	19	11	17
5 (D3EF)	11 (D3R)	19	8	9
5 (D3EF)	12 (M3R)	19	9	10
			Covariance = 5.36	
G				
6 (M3EF)	7 (D1R)	18	12	16
6 (M3EF)	8 (M1R)	18	10	16
6 (M3EF)	9 (D2R)	18	11	17
6 (M3EF)	10 (M2R)	18	12	17
6 (M3EF)	11 (D3R)	18	8	9
6 (M3EF)	12 (M3R)	18	9	10
			Covariance = 4.4	

Table 3. Comparison of selected sample characteristics, numbers of fish species captured in given samples (or sample pools) and species richness estimators' values. *CT*: number of species captured by DSN (in site 1) or EF (sites 2 and 3) in a sample (or a sample pool); R: number of species captured by the control, i.e. rotenone; *CT + R*: number of species captured by both DSN and rotenone (site 1) or EF and rotenone (sites 2 and 3) in a sample (or a sample pool); *R* only: number of species captured by rotenone but not by DSN or EF in a given sample (or a sample pool); HM through G-P CMLE: species richness estimators' values (± their S.E.) calculated on the basis of abundance data (pooled in the case of sample combinations) of those species that were captured by DSN or EF (i.e. those in column *CT*).

Samples	Site 1	Site 2	Site 3	Dec	Mar	CT	R	CT+R	R only	HM	HM (MLE)	Chao 1	ACE	ACE 1
1	+			+		12	16	19	7	13.1 ± 1.5	12.0 ± 0.0	13.5 ± 2.3	16.3 ± 4.7	18.6 ± 8.3
2	+				+	12	16	20	8	30.7 ± 21.5	12.0 ± 0.0	33.0 ± 17.3	48.8 ± 46.4	77.0 ± 94.5
1+2	+			+	+	16	22	26	10	19.9 ± 3.5	16.0 ± 0.0	22.0 ± 6.5	28.0 ± 10.1	36.9 ± 21.7
3		+		+		19	17	21	2	19.7 ± 1.0	19.0 ± 0.0	22.0 ± 4.6	20.7 ± 2.2	21.1 ± 2.8
4		+			+	21	17	23	2	22.2 ± 1.3	21.0 ± 0.0	23.0 ± 2.6	23.8 ± 2.7	24.3 ± 3.4
5			+	+		19	9	20	1	19.3 ± 0.6	19.0 ± 0.0	19.2 ± 0.5	19.4 ± 0.8	19.4 ± 0.8
6			+		+	18	10	19	1	23.1 ± 4.5	18.0 ± 0.0	30.3 ± 13.2	33.7 ± 13.6	47.2 ± 31.5
3+4		+		+	+	26	24	28	2	27.3 ± 1.4	26.0 ± 0.0	30.2 ± 4.9	30.2 ± 3.8	31.6 ± 5.6
3+5		+	+	+		23	21	25	2	23.6 ± 0.9	23.0 ± 0.0	25.0 ± 3.7	24.1 ± 1.6	24.3 ± 1.9
3+6		+	+	+	+	25	20	26	1	27.9 ± 2.5	25.0 ± 0.0	53.0 ± 21.4	36.3 ± 9.4	45.3 ± 20.8
4+5		+	+	+	+	28	21	30	2	29.7 ± 1.7	28.0 ± 0.0	30.5 ± 3.0	32.2 ± 3.7	33.3 ± 5.0
4+6		+	+		+	24	20	27	3	26.7 ± 2.3	24.0 ± 0.0	45.0 ± 17.4	31.6 ± 6.2	35.0 ± 10.4
5+6			+	+	+	24	15	25	1	25.6 ± 1.6	24.0 ± 0.0	36.5 ± 17.1	28.2 ± 3.9	29.5 ± 5.8
3+4+5		+	+	+	+	30	26	32	2	31.2 ± 1.4	30.0 ± 0.0	32.0 ± 2.6	32.8 ± 2.8	33.4 ± 3.5
3+4+6		+	+	+	+	29	25	31	2	31.7 ± 2.2	29.0 ± 0.0	61.0 ± 39.6	40.1 ± 8.2	47.6 ± 16.6
3+5+6		+	+	+	+	27	22	28	1	28.6 ± 1.6	27.0 ± 0.0	37.0 ± 10.3	30.8 ± 3.7	31.9 ± 5.2
4+5+6		+	+	+	+	29	22	31	2	31.2 ± 2.0	29.0 ± 0.0	47.0 ± 23.6	33.7 ± 4.1	34.9 ± 5.6
3+4+5+6		+	+	+	+	31	26	33	2	32.7 ± 1.6	31.0 ± 0.0	43.5 ± 17.1	34.4 ± 3.2	35.0 ± 4.1

Samples	Site 1	Site 2	Site 3	Dec	Mar	CT	R	CT+R	R only	Jackknife 1	Jackknife 2	G-P M	G-P UMLE	G-P CMLE
1	+			+		12	16	19	7	15.0 ± 2.4	15.0 ± 4.2	22.5 ± 24.0	16.7 ± 10.0	* *
2	+				+	12	16	20	8	18.9 ± 3.7	25.8 ± 6.4	* *	* *	* *
1+2	+			+	+	16	22	26	10	22.0 ± 3.5	25.0 ± 6.0	* *	* *	* *
3		+		+		19	17	21	2	22.0 ± 2.4	25.0 ± 4.2	21.2 ± 3.9	19.0 ± 1.1	* *
4		+			+	21	17	23	2	25.0 ± 2.8	25.0 ± 4.9	24.4 ± 4.2	22.9 ± 2.9	24.8 ± 7.0
5			+	+		19	9	20	1	20.0 ± 1.4	18.0 ± 2.4	19.4 ± 2.3	* *	19.5 ± 4.9
6			+		+	18	10	19	1	25.0 ± 3.7	30.0 ± 6.5	* *	* *	* *
3+4		+		+	+	26	24	28	2	31.0 ± 3.2	33.0 ± 5.5	32.7 ± 8.6	30.2 ± 6.1	* *
3+5		+	+	+		23	21	25	2	25.0 ± 2.0	26.0 ± 3.5	24.2 ± 3.1	22.9 ± 1.0	24.0 ± 4.7
3+6		+	+	+	+	25	20	26	1	33.0 ± 4.0	41.0 ± 6.9	* *	* *	* *
4+5		+	+	+	+	28	21	30	2	33.0 ± 3.2	33.0 ± 5.5	33.8 ± 6.8	32.2 ± 5.4	35.3 ± 16.1
4+6		+	+		+	24	20	27	3	31.0 ± 3.7	38.0 ± 6.5	41.2 ± 27.9	30.7 ± 9.6	* *
5+6			+	+	+	24	15	25	1	29.0 ± 3.2	33.0 ± 5.5	30.4 ± 8.9	26.3 ± 3.7	* *
3+4+5		+	+	+	+	30	26	32	2	34.0 ± 2.8	34.0 ± 4.9	33.4 ± 4.7	32.0 ± 2.9	33.8 ± 6.7
3+4+6		+	+	+	+	29	25	31	2	37.0 ± 4.0	44.0 ± 6.9	101 ± 269	* *	* *
3+5+6		+	+	+	+	27	22	28	1	32.0 ± 3.2	37.0 ± 5.5	32.6 ± 7.8	28.5 ± 2.7	30.6 ± 11.6
4+5+6		+	+	+	+	29	22	31	2	35.0 ± 3.5	40.0 ± 6.0	35.5 ± 7.9	31.8 ± 3.9	34.2 ± 10.5
3+4+5+6		+	+	+	+	31	26	33	2	36.0 ± 3.2	40.0 ± 5.5	35.2 ± 5.6	32.5 ± 2.4	34.1 ± 5.6

to test how the samples obtained affected 10 species richness estimators (Chao and Shen 2003) (Tables 3 and 4). The samples that were used for this analysis were those obtained before the application of rotenone, i.e. they were either DSN or EF samples (samples 1 through 6). Data input files were either separate samples or samples pooled in various ways. This was done to observe how the species richness estimators performed on cumulated and heterogeneous data sets. Because EF proved much more efficient in sampling then DSN all EF sample combinations are included (Table 3). As expected, DSN samples caused much variation in species richness estimators (Table 3). Combining them with EF samples reduced the variation but as this was produced by EF they gave little insight into their nature and the respective pools are not included.

Table 4. Part A: Spearman *r* correlation coefficient values between the number of species actually captured by EF and rotenone (*CT* + *R*) and numbers of species predicted (on the basis of EF alone) by the species richness estimators of Table 3; P is significance level. Part B: percentage divergences of given estimators' values from *CT* + *R* values (in Table 3) for given samples (or sample pools). Part C: mean precision, i.e. the means of percentage divergences in Part B. Part D: bias of estimates in Part B, i.e. ratio of overestimates to underestimates.

		HM	HM (MLE)	Chao 1	ACE	ACE 1	Jackknife 1	Jackknife 2	G-P M	G-P UMLE	G-P CMLE
A	CT+R	0.96	0.99	0.64	0.62	0.46	0.92	0.74	0.37	0.20	−0.01
	P	0.00	0.00	0.01	0.01	0.08	0.00	0.00	0.17	0.49	0.96
B	3	−6.6	−10.5	4.5	−1.4	0.5	4.5	16.0	0.9	−10.5	*
	4	−3.6	−9.5	0.0	3.4	5.3	8.0	8.0	5.7	−0.4	7.3
	5	−3.6	−5.3	-4.2	−3.1	−3.1	0.0	−11.1	−3.1	*	−2.6
	6	17.7	−5.6	37.3	43.6	59.7	24.0	36.7	*	*	*
	3+4	−2.6	−7.7	7.3	7.3	11.4	9.7	15.2	14.4	7.3	*
	3+5	−5.9	−8.7	0.0	−3.7	−2.9	0.0	3.8	−3.3	−9.2	−4.2
	3+6	6.8	−4.0	50.9	28.4	42.6	21.2	36.6	*	*	*
	4+5	−1.0	−7.1	1.6	6.8	9.9	9.1	9.1	11.2	6.8	15.0
	4+6	−1.1	−12.5	40.0	14.6	22.9	12.9	28.9	34.5	12.1	*
	5+6	2.3	−4.2	31.5	11.3	15.3	13.8	24.2	17.8	4.9	*
	3+4+5	−2.6	−6.7	0.0	2.4	4.2	5.9	5.9	4.2	0.0	5.3
	3+4+6	2.2	−6.9	49.2	22.7	34.9	16.2	29.5	69.2	*	*
	3+5+6	2.1	−3.7	24.3	9.1	12.2	12.5	24.3	14.1	1.8	8.5
	4+5+6	0.6	−6.9	34.0	8.0	11.2	11.4	22.5	12.7	2.5	9.4
	3+4+5+6	−0.9	−6.5	24.1	4.1	5.7	8.3	17.5	6.3	−1.5	3.2
C	Mean	0.3	−7.0	20.0	10.2	15.3	10.5	17.8	12.3	0.9	2.8
D	Bias	6/9	0/15	14/1	12/3	13/2	15/0	14/1	11/2	7/4	6/2

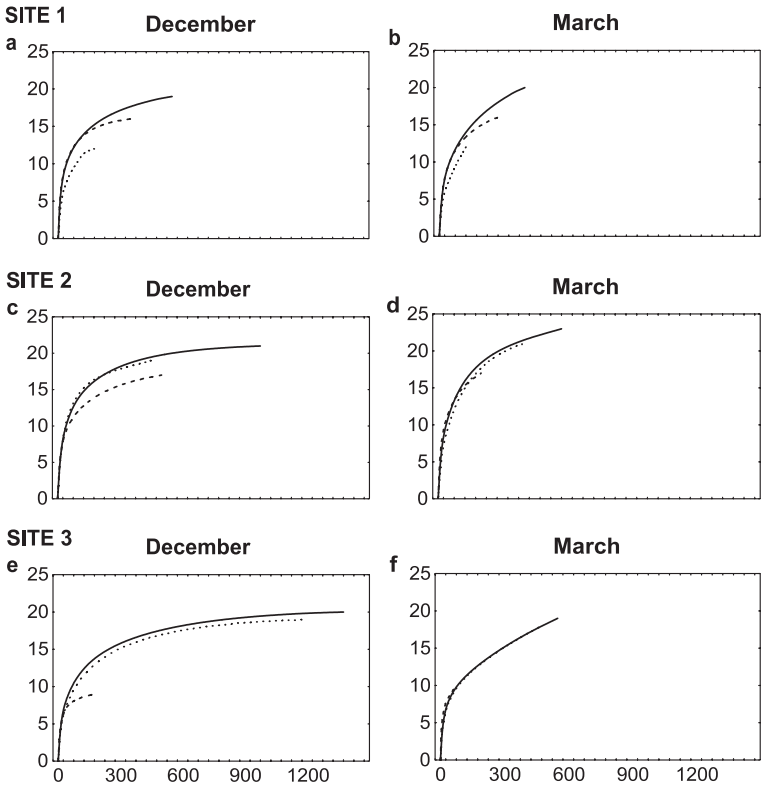

Fig. 1. Rarefaction curves for samples collected in the Taquaral River compared by sites. Dotted line indicates Double stick net, DSN (site 1) or Electrofishing, EF samples (sites 2 and 3), dashed line rotenone samples, and solid line DSN + rotenone (site 1) or EF + rotenone samples (sites 2 and 3). *Y*- vertical axis: number of fish species; *X*- horizontal axis: number of specimens.

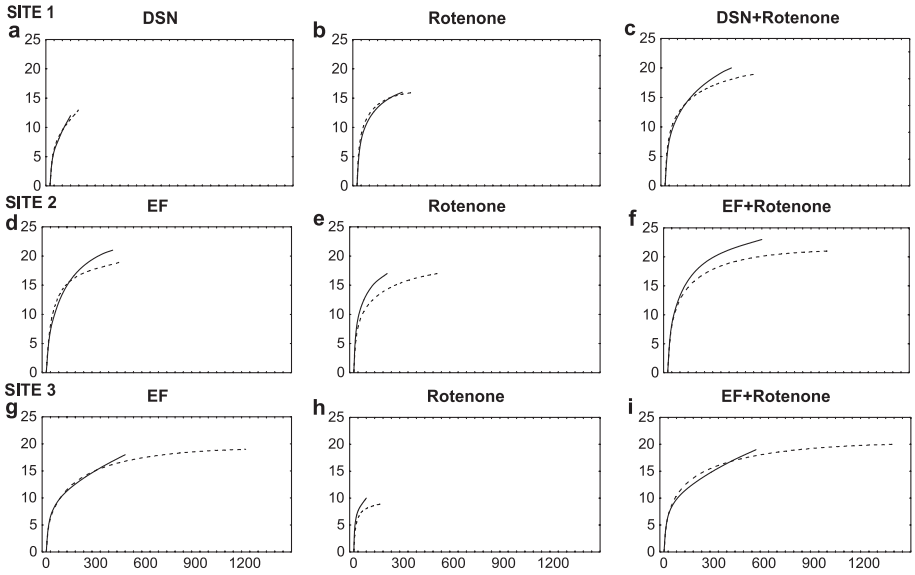

Fig. 2. Rarefaction curves for samples collected in the Taquaral River compared by seasons. Dashed line indicates December samples, solid line indicates March samples. *Y*- vertical axis: number of fish species; *X*- horizontal axis: number of specimens.

4 Discussion

Fishing gears are physically calibrated either by repetitive usage (i.e. several sampling runs or passes) of the same technique (Meador et al. 2003, for example, used two-pass backpack electrofishing to calibrate single-pass backpack electrofishing, while Reynolds et al. 2003 a two-pass one with a third pass), or by application of various gears, or by combinations of both (Casselman et al. 1990).

The underlying assumption in calibrating fishing gears with one another is that some are more efficient or/and differently selective that others. In Penczak et al.'s (2003) study rotenone applied after DSN and EF was used to calibrate them in the Brazilian stream, assuming that rotenone was more efficient than the two other gears; in fact, that this combination will capture all species. This was a plausible assumption, considering, for example, Fisher (1987) who assumed that "sodium cyanide plus electrofishing [with which he calibrated a quadrat sampler] was 100% effective".

Fishing gears may also be calibrated by indirect, extrapolation methods, such as species richness estimators and heterogeneity indices (Burnham and Overton 1978; Colwell and Coddington 1994; Chao 1984; Chao 1987; Chao and Lee 1992; Chao et al. 2000; Chao and Shen 2004; Chao et al. 2005). The advantage of these is that they are data independent, hence the danger of constructing models with the same data that serve for the prediction of the models' efficiency is avoided (Olden et al. 2002).

The simplest of such methods is analysis of sample pairs and their common species (Table 2). It was of limited use. The explanation of the strong difference in covariance between the numbers of common species and species captured by single EF samples and their respective rotenone samples (Table 2A) is actually easy to explain: several species that were missing in the rotenone sample had already been depleted by EF. Thus, probably, this difference has nothing to do with selectivity of the gears. The comparison of covariance between given EF samples and all the rotenone samples (Tables 2D-G) revealed much difference between site 2 and site 3, yet it is difficult to determine whether this difference resulted from actual difference in the quality of the sites or from other factors.

In general, little evidence for EF selectivity was detected. Ca. 80–90% of species constituting DSN samples were captured by EF and about 60–70% by rotenone. Had there been strong between-site or between-season selectivity we would have obtained little overlap in species captured by gears in different sites and/or different seasons.

The most surprising thing concerning the EF samples is that although the species captured by EF and/or rotenone much varied from sample to sample and gradually increased in number together with pooling the samples, a number of species that was added by the control (i.e. rotenone) to that of an EF sample was almost constant (Table 3, column "*R* only"), in other words column *CT* + *R* differed from column *CT* of Table 3 by 1 or 2 species (3 in only one case). Note that this occurred irrespectively of considerable diversity in several parameters of the Taquaral data (Table 5). It is hard to deduce any explanation for this phenomenon. One thing it might suggest is that the efficiency of both conventional gears (i.e. EF and rotenone) was independent of any such factors as the structure of the fauna, biology of the fish species, stages of reproduction cycles, environmental conditions, etc. The opposite solution would be that both gears are identically biased, but in view of the diversity in all the above factors this is very unlikely. Consequently, rotenone would be a very good estimator of the samples – unfortunately, we do not know whether its efficiency is similar in any other region and/or scale than in the Brazilian stream. Worse still, this can hardly be tested as rotenone application is forbidden in most countries.

Some of the species richness estimators applied in the present study were also applied to fish by several other authors. Hughes et al. (2002) applied ACE, Chao 1, Jackknife 1 and Jackknife 2 to several dozen streams of Oregon, USA, to

Table 5. Spearman correlation coefficients and their respective significance levels calculated between selected variables of Table 3 (in all cases $n = 15$).

	Difference between columns *CT* and *CT* + *R* of Table 3
Number of samples pooled (obtained from column "Samples" of Table 3)	$r = 0.217$, $p = 0.437$
Number of sites (obtained from columns "Site 2" and "Site 3" of Table 3)	$r = 0.327$, $p = 0.234$
Number of seasons (obtained from columns "Dec" and "Mar" of Table 3)	$r = -0.109$, $p = 0.699$
Number of EF captured species (column "*CT*" of Table 3)	$r = 0.316$, $p = 0.251$
Number of rotenone captured species (column "*R*" of Table 3)	$r = 0.462$, $p = 0.083$
Number of EF and rotenone captured species (column "*CT* + *R*" of Table 3)	$r = 0.434$, $p = 0.106$

each reach they studied. The performance varied. Occasionally, they over- or underestimated the species numbers actually captured by several dozen percent, but Hughes et al. (2002) did not consider any of the predictors as more or less correct than others.

Palmer (1991) also suggested evaluating richness estimators with simulated data beside real data. This was in a way accomplished by Olden et al. (2002), who studied fish in lakes and rivers of Ontario and British Columbia, Canada. They noticed that Jackknife type of approach to species distribution models is superior to validation based on resubstitution approach (i.e. when the same data are used for both model construction and prediction of its efficiency).

In the present study, the reliability of estimators much varied as regards given DSN samples (Table 3). Most of them were consistent in this respect. The bias was either crude underestimation (as in sample 1), except G-P M and ACE 1, which produced higher values, or overestimation (as in sample 2), except Jackknife 1, which produced a lower one. Pooling of both these samples alone positively affected the stability and correctness of the estimators' predictions and they performed best when applied to the total pool (Table 3). This seems to indicate that the samples were strongly seasonally biased (as they were from different seasons).

A different image was produced by EF samples (Tables 3, 4 and 5). One general observation is that in the case of some samples and their pools G-P M, G-P UMLE and GP CMLE produced no estimates. The other is that the estimators varied both from one another and in terms of samples (or pools) that served as the basis for prediction.

The first estimator, HM, was very accurate in its prediction. This concerned all estimates of sample pools and of all single samples, except only number 6, which was much overestimated. Despite the impact of sample 6 HM most highly (except HM (MLE)) correlated with the actually captured species numbers (Table 4). To employ Palmer's (1990, 1991) terminology HM was very precise as the mean percentage difference from actual samples was 0.3% (Table 4); this value was the lowest of all the estimators and would make it a very good one of the Taquaral fish fauna had it not been for the value of sample 6. HM was also the least biased, as the ratio of overestimates to underestimates was 6/9. Note another specific feature of HM, namely its increasing precision together with increasing sample pooling (while most estimators behaved otherwise), and note the fact that although sample 6 alone negatively affected the estimator, its impact did not proliferate to those pools that included sample 6, as was mostly the case in other estimators. Finally, HM displayed very narrow S.D. ranges as compared with other estimators (Table 3).

Another good estimator turned out Jackknife 1. It correlated almost as highly with the actually captured number as HM, yet contrary to the latter was much biased and little precise (Table 4). Note that its precision was lower than even that of weakly correlated ACE. Similarly as HM it was strongly negatively affected by sample 6, yet contrary to HM this impact also proliferated to those pools that contained sample 6.

Much the same may be concluded as regards Jackknife 2. However, it differed from the former by being even stronger affected by sample 6 (including all its pools) and by the fact that it was much more incorrect in the case of all single samples except 4. Jackknife 2 displayed the highest overestimating bias, and its average precision and correlation with the actually captured species numbers were low (Table 4). It also displayed the highest single underestimate of all estimates (in sample 5) (ignoring that of HM (MLE)).

Chao 1 was even stronger than Jackknife 2 affected by sample 6. In fact, it seems that in this estimator the impact is most conspicuous, because the pools with sample 6 are most overestimated, and also because other single samples (than 6) and their pools were quite correctly estimated. It was almost as positively biased as Jackknife 1, while its precision and correlation were lower (Table 4).

In both ACE and ACE 1, sample 6 (considered in isolation) produced the highest bias of all the estimators. Yet, contrary to particularly Chao 1, this impact decreased with increase in sample pooling and in the last sample combination (3+4+5+6) was very low. ACE and ACE 1 were also quite correct with single samples (except sample 6). They were less biased and more precise than Chao 1, yet worse correlated with the actually captured species numbers (Table 4).

Less may be concluded as regards G-P M, G-P UMLE and G-P CMLE as they produced no estimates in 2, 4 and 7 cases, respectively. None did calculate sample 6 alone. G-P M also produced the highest single overestimate of all the estimators (69.2% for pool 3+4+5). Although G-P UMLE and G-P CMLE displayed very high mean precision yet they worst correlated with actual numbers of species. G-P M correlated little better, while being highly positively biased.

A case apart is HM(MLE). It most highly correlated with the actual species numbers of all the estimators (0.99, Table 4),

yet it only reported the numbers that were actually captured by EF, producing in fact no estimates at all (so its "estimates" are the same numbers as those in column *CT* of Table 3). Its poor "estimation" is manifest in its most negative bias and in its negative precision (no other estimator produced mean negative precision), which is also considerable.

Looking at the estimates from the sample/pool perspective it is evident that samples 5 and 6 were most specific as the former produced under- and the latter overestimation (ignoring HM(MLE) of course) in all the estimators. One reason of this different impact of samples 5 and 6 might be their seasonal difference: the first was obtained in December, the other in March. In the southern hemisphere March is an advanced stage of the growing season while December is its middle (although in the subtropical zone seasonal differences are not as clear cut as in the temperate one). Consequently, it might be presumed that the March sample is more stable than the December one, when new cohorts are spawn. Thus, it seems that the estimators might be least biased in the case of assemblages that are in some intermediate stage between December and March.

Note that such over- and underestimation did not occur in the case of samples 3 and 4, which are also of contrasting seasons. These two samples, however, were collected in a site that differed from that of samples 5 and 6 in being much more overgrown by vegetation and much less stony. Consequently, perhaps this environmental factor "neutralized" the impact of seasonality. This impact of site difference was manifest both in the case of species richness and species abundance (i.e. number of individuals) as regards the proportion of EF/R efficiency. As regards number of individuals rotenone was always much less efficient than a (combined) EF sample, except site 2 in December (the respective difference measured by the Chi-square test (2×2 table) was always significant at $p = 0.000$), no matter whether the EF/R proportion was compared within and/or between sites, and within and/or between seasons. But in each season the proportion also differed statistically between sites (same test and same p level). This was different when EF/R efficiency was compared in terms of species richness (Chi-square and Fisher exact test values were $p = 0.17$ and $p = 0.11$, when respective values of both seasons were combined in a 2×2 table, for example). Please remember, however, that the lower (quantitative and qualitative) efficiency of rotenone resulted from fish having been depleted by EF, particularly in site 3; if rotenone was used independently it would certainly be more efficient as the last EF subsample of each sample (D3EF4, for example) was always less abundant than the rotenone one that followed it.

There are two other environmental factors that may also affect the structure and richness of fish fauna and thus of species richness estimates. One is water velocity. As it was found by Peterson and Rabeni (2001) fish families display high differences in catchability (these ranged from 84% in Cyprinidae to 31% in Ictaluridae in their study). As velocity in sites 1 and 2 of the Brazilian stream was much higher in March than in December the ratio of species belonging to given families may have changed, as for cyprinids, for example, (Peterson and Rabeni 2001) it is much more difficult to stay in high current, while some Cottidae have morphological adaptations to withstand strong currents (Pflieger 1997). Also Ictaluridae and Percidae tend to hide in interstitial spaces in coarse gravel or beneath large cobbles (Pflieger 1997), which makes them difficult to capture. However, these families did not occur in our data and there is no data on whether water velocity may be selective for characids (Nelson 1994), which dominated the samples. Other families that occur in the Taquaral, like silurids, are typical reophilic species, so this factor does not apply to them as well. In the Taquaral, velocity was considerably lower in site 2 in December, which co-occurred with a significantly different EF/R efficiency measured in terms of captured individuals (Chi-square of $p = 0.000$), but the proportion of captured species was not significantly different then.

The other factor are canopy cover and substratum. Sites 1 and 2 were strongly canopied with branches overhanging into the water and there were snags in the current. Consequently, the sites may have possessed more hiding places of vegetation origin than site 3 and the canopy must have affected illumination of the sites (no water temperature changes were recorded). On the other hand sites 1 and 2 possessed much less stones, which must have also constituted hiding places in site 3.

It is hard to determine how the above factor may have impacted the fish, and the data we obtained do not seem to have been affected by any of the above factors, particularly by making our catches species-selective. The other thing is that rotenone added different species to each electrofished sample, and pooling of samples did not disturb the pattern: no rotenone-captured species occurred twice when heterogeneous sites (samples 3+5, 4+6) or seasons (samples 3+4, 5+6) were compared, for example. Even a comparison of samples obtained by different gears was congruent with the pattern as there was only one species that co-occurred (in two samples) when samples were compared within seasons (sample 2+4+6).

Of course, we cannot determine from the scarce data we have managed to obtain whether the presently reported performance of the species richness estimators is typical in any other scale than that of fish faunas of small tropical Brazilian streams, not to mention other animal groups. Similarly, until richer data are collected we cannot be sure if EF can be as successful a tool in predicting species richness in other situations.

To sum up, rotenone decisively turned out the best sampling method in the subtropical small stream. However, electrofishing combined with the HM estimator was quite satisfactory for reliable qualitative (species richness) predictions when at least several sampled were pooled. This is important as the use of rotenone is almost completely forbidden worldwide, small nets are little efficient and very selective, while large ones are impossible to use in fast flowing streams, even quite small, due to the force of water.

Acknowledgements. We are very obliged to Professor Robert Colwell for his comment on our application of EstimateS software. Especial thanks are directed to Professor Anne Chao for giving us access to her SPADE software and for precious and extensive suggestions on how to apply it for analysis. Furnas Centrais Electricas and Nupelia (Maringá State University) are thanked for the financial and people support for sampling. Especial thanks are directed to Prof. AA Agostinho for organizing the field research. The field investigations on which this study is based complied with the current laws of the country in which they were performed.

References

Bayley P.B., Peterson J.T., 2001, An approach to estimate probability of presence and richness of fish species. Trans. Am. Fish. Soc. 130, 620-633.

Burnham K.P., Overton W.S., 1978, Estimation of the size of a closed population when capture probabilities vary among animals. Biometrika 65, 623-633.

Casselman J.M., Penczak T., Carl L., Mann R.H.K., Holcik J., Woitowich W.A., 1990, An evaluation of fish sampling methodologies for large river systems. Pol. Arch. Hydrobiol. 37, 521-551.

Chao A., 1984, Non-parametric estimation of the number of classes in a population. Scand. J. Stat. 11, 265-270.

Chao A., 1987, Estimating the population size for capture-recapture data with unequal catchability. Biometrics 42, 783-791.

Chao A., Bunge J., 2002, Estimating the number of species in a stochastic abundance model. Biometrics 58, 531-539.

Chao A., Lee S.-M., 1992, Estimating the number of classes via sample coverage. J. Am. Stat. Assoc. 87, 210-217.

Chao A., Shen T.-J., 2003, Program SPADE (Species Richness And Diversity Estimation). Program and User's Guide published at `http://chao.stat.nthu.edu.tw` (Program latest updated Feb 2005, User Guide latest updated Apr 2004).

Chao A., Shen T.-J., 2004, Nonparametric prediction in species sampling. J. Agric. Biol. Environ. S. 9, 253-269.

Chao A., Hwang W.-H., Chen Y-C., Kuo C.-Y., 2000, Estimating the number of shared species in two communities. Stat. Sinica 10, 227-246.

Chao A., Chazdon R.L., Colwell R.K., Shen T.-J., 2005, A new statistical approach for assessing similarity of species composition with incidence and abundance data. Ecol. Lett. 8, 148-159.

Colwell R.K., 1997, EstimateS: Statistical estimation of species richness and shared species from samples. Version 5. User's Guide and application. Published at: `http://viceroy.eeb.uconn.edu/estimates`.

Colwell R.K., Coddington J.A., 1994, Estimating terrestrial biodiversity through extrapolation. Phil. Trans. R. Soc. B 345, 101-118.

Davies W.D., Shelton W.L., 1983, Sampling with toxicant. In: Nielsen L.A., Johnson D.L., Lampton S.S. (Eds.) Fisheries techniques. American Fisheries Society, Bethesda, pp. 199-213.

Fisher W.L., 1987, Benthic fish sampler for use in riffle habitats. Trans. Am. Fish. Soc. 116, 768-772.

Głowacki Ł., Penczak T., 2000, Impoundment impact on fish in the Warta River: species richness and sample size in the rarefaction method. J. Fish Biol. 57, 99-108.

Heck K.L., Jr., van Belle G., Simberloff D., 1975, Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. Ecology 56, 1459-1461.

Hughes R.M., Kaufmann P.R., Herlihy A.T., Intelmann S.S., Corbett S.C., Arbogast M.C., Hjort R.C., 2002, Electrofishing distance needed to estimate fish species richness in raftable Oregon rivers. N. Am. J. Fish. Manage. 22, 1229-1240.

Hurlbert S.H., 1971, The nonconcept of species diversity: A critique and alternative parameters. Ecology 52, 577-586.

Jackson D.A., Peres-Neto P.R., Olden J.D., 2001, What controls who is where in freshwater fish communities – the role of biotic, abiotic and spatial factors. Can. J. Fish. Aquat. Sci. 58, 157-170.

Ludwig J.A., Reynolds J.F., 1988, Statistical Ecology. John Wiley, New York.

Magurran A.E., 1988, Ecological diversity and its measurement. Princeton University Press, London.

Matthews W.J., 1998, Patterns in freshwater fish ecology. Chapman & Hall, and International Thomson Publishing, New York.

Meador M.R., McIntyre J.P., Pollock, K.H., 2003, Assessing the efficacy of single-pass backpack electrofishing to characterize fish community structure. Trans. Am. Fish. Soc. 132, 39-46.

Nelson J.S., 1994, Fishes of the world. Third edition. John Wiley & Sons, Inc., New York.

Oberdorff T., Guégan J.F., Hugueny B., 1995, Global scale patterns of fish species richness in rivers. Ecography 18, 345-352.

Olden J.D., Jackson D.A., Peres-Neto P.R., 2002, Predictive models of fish species distributions: A note on proper validation and chance predictions. Trans. Am. Fish. Soc. 131, 329-336.

Palmer M.W., 1990, The estimation of species richness by extrapolation. Ecology 71, 1195-1198.

Palmer M.W., 1991, Estimating species richness: The second-order jackknife reconsidered. Ecology 72, 1512-1513.

Palmer M.W., 1994, Variation in species richness: Towards a unification of hypotheses. Folia Geobot. Phytotax., Praha 29, 511-530.

Palmer M.W., 1995, How should one count species? Nat. Area J. 15, 124-135.

Peet R.K., 1974, The measurement of species diversity. Ann. Rev. Ecol. Syst. 5, 285-307.

Penczak T., Agostinho A.A., Latini J.D., 2003, Rotenone calibration of fish density and standing crop in a tropical stream estimated by two removal methods. Hydrobiologia 510, 23-38.

Penczak T., Gomes L.C., Bini L.M., Agostinho A.A., 1998, The importance of qualitative inventory sampling using electric fishing and nets in a large, tropical river (Brazil). Hydrobiologia 389, 89-100.

Penczak T., Jakubowski H., 1990, Drawbacks of electric fishing in rivers. In: Cowx I. G. (Ed.) Developments in Electric Fishing. Fishing News Books, Oxford, pp. 115-122.

Peterson J.T., Rabeni C.F., 2001, Evaluating the efficiency of a one-square-meter quadrat sampler for riffle-dwelling fish. N. Am. J. Fish. Manage. 21, 76-85.

Pflieger W.L., 1997, The fishes of Missouri (revised). Missouri Department of Conservation, Jefferson City.

Reynolds L., Herlihy A.T., Kaufmann P.R., Gregory S.V., Hughes R.H., 2003, Electrofishing effort requirements for assessing species richness and biotic integrity in Western Oregon streams. N. Am. J. Fish. Manage. 23, 450-461.

Shen T.-J., Chao A., Lin C.-F., 2003, Predicting the number of new species in further taxonomic sampling. Ecology 84, 798-804.

StatSoft, Inc., 2000, Statistica for Windows ver. 5.5 A PL. Tulsa, OK, USA, `http://www.statsoft.com.pl`.